Nugroho. (2024)
https://doi.org/10.70310/7y9mnh17

ORIGINAL ARTICLE Open Access

Data Science: Building Visual Understanding

Agung Yuliyanto Nugroho[1]

Abstract
Data science is rapidly evolving and plays a vital role in a variety of fields, from business to healthcare. However, the complexity of the data generated often makes it difficult for stakeholders without a technical background to understand. Data visualization is here to bridge this gap, enabling complex information to be translated into more intuitive and understandable graphical representations. Through effective visualization, patterns, trends, and insights hidden in data can be identified more quickly, facilitating better and faster decision-making. This article explores the critical role of visualization in data science, discussing the key techniques used to transform raw data into informative visualizations. It will also discuss how choosing the right visualization can help uncover relationships between variables, facilitate predictive analysis, and support in-depth analytical processes. By integrating visualization into the data science workflow, we can not only accelerate understanding but also improve communication between technical and non-technical teams. Ultimately, visualization is not just a supporting tool, but a core element that helps transform data into valuable insights.

Keywords *Data Visualization, Data Science, Graphical Representation, Data Analysis*

Introduction

In the era of increasingly developing information, the amount of data generated every day continues to increase exponentially. This data covers various sectors, ranging from business, health, education, to government, which produce large amounts of data and in complex forms. To utilize this data effectively, a method is needed that is able to simplify and present data in a form that is easier to understand. This is where the role of data visualization becomes very important.

The process of converting raw data into graphical representations such as graphs, diagrams, maps, and various other visual forms. This process not only helps in understanding the information hidden in the data, but also accelerates decision making, both in the context of business, research, and public policy. Effective data visualization allows users, including those without a technical background, to identify patterns, trends, and anomalies in data intuitively.

However, although data visualization has been widely used, the main challenge faced in this study is how to create appropriate and effective visualizations. Not all visualizations can convey information well; mistakes in choosing the type of graph or inappropriate presentation can actually lead to misinterpretation of the data. Therefore, a systematic approach is needed to determine the most appropriate visualization method for the type of data and the purpose of the analysis.

This study aims to examine the important role of data visualization in the data analysis process, as well as identify the most effective visualization techniques for various types of data.

*Koresponden:
Agung Yuliyanto Nugroho
agungboiler11@gmail.com
Informasi daftar penulis tersedia di bagian akhir artikel

© The Author(s) 2024. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Thus, this study is expected to provide better guidance in selecting and developing visualizations that are not only informative, but also able to simplify the complexity of data so that it can be utilized optimally.

Why Data Visualization is Important because it Facilitates Data Understanding: Visualization helps us see patterns, trends, and anomalies that may not be visible in raw data. By looking at a graph or map, one can quickly gain the necessary insights.

Helps Communication: One of the main goals of data science is to convey findings to stakeholders. Data visualization helps communicate the results of data analysis in a way that is easy to understand, especially for those who do not have a technical background.

Identify Patterns and Trends: Visualization allows us to easily identify relationships between variables, historical trends, and recurring patterns that may be difficult to find by just looking at the numbers.

Decision Making: By understanding data through visualization, decision makers can make more informed and informed decisions based on the data presented.

1. Bar and Column Charts: Used to compare categories or numbers in a dataset. Line Charts: Used to see data trends over time. Pie Charts: Used to see distributions or proportions between categories.
2. Heatmaps: Used to display correlations between variables in the form of a color matrix.
3. Histograms: Present the frequency distribution of numeric data.

Box Plots: Useful for understanding the distribution, outliers, and variability of data.

Geographic Maps: Useful for presenting spatial data, for example to see demographic distributions or disease patterns.

Some popular tools that are often used to create data visualizations in data science include:

Matplotlib and Seaborn: Very popular visualization libraries in Python for creating static graphs.

Tableau: An interactive data visualization tool often used to present dashboards.

Power BI: Software from Microsoft that provides a platform for data analysis and visualization creation.

Literature Review

According to Ware (2004), data visualization is a graphical representation of data that aims to make it easier for users to understand, interpret, and filter information. Tufte (2001) also emphasized the importance of effective visualization to convey information efficiently without sacrificing data accuracy. This study highlights that visualization not only makes data easier to understand, but also serves as a tool to explore and analyze data.

Data Visualization

Several theories have been developed to understand how humans process visual information. The Graphic Theory by Bertin (1983) suggests that visual elements such as position, size, shape, and color are basic tools in presenting information graphically. On the other hand, the Cognitive Mapping Theory by Card, Mackinlay, and Shneiderman (1999) examines how visualization helps

users connect various data elements through visual representations, which facilitates the understanding of patterns and correlations.

Types of Data Visualization

A study by Few (2013) classifies data visualization into several main types such as bar graphs, line graphs, pie charts, heat maps, and geographic maps. This study emphasizes that the selection of the right type of visualization must be adjusted to the characteristics of the data presented and the analysis objectives to be achieved. For example, line graphs are more effective in showing changes over time, while heat maps are suitable for showing data distribution in a more compact space.

Modern Visualization Tools

The development of information technology also enriches modern visualization techniques. Tableau, Power BI, and D3.js are examples of tools that enable integration between data analysis and visual representation. According to Healy (2018), these tools provide users with the flexibility to create interactive visualizations, which allow for deeper exploration of larger and more complex data. Interactive tools like these strengthen users' ability to perform dynamic analysis compared to static visualizations.

Quality and Effectiveness of Visualization

The quality of data visualization is measured by its ability to convey information accurately and efficiently. Cleveland and McGill (1985) suggested that effective visualization should minimize data distortion and maximize understanding. Research by Heer et al. (2010) also added that the effectiveness of visualization is influenced by how well users can interact with the data and gain useful insights from the visualization.

Challenges in Data Visualization

Although data visualization has many benefits, there are some challenges that need to be considered. One of the main challenges is the misperception that can occur if the visualization is not well designed (Kirk, 2016). Mistakes in choosing scale, color, or shape can cause bias in data interpretation. In addition, the limitations of visualization tools in handling very large data (big data) are also a concern discussed in the literature by Keim et al. (2013).

Visualization in Decision Making

A study by McCandless (2010) emphasized that data visualization can significantly support decision-making in business and government contexts. Good visualization allows decision-makers to see patterns and trends that are not visible through tables or numerical reports, thus accelerating the data-driven decision-making process.

Metode

This study uses a descriptive-analytical method, which aims to explore and understand the role of data visualization in the data analysis process. Through this method, the study will describe

various visualization techniques used in data science practices and evaluate their effectiveness in presenting information clearly and accurately.

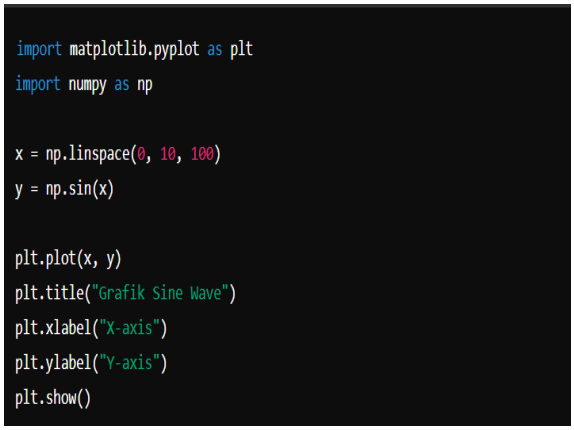
Results and Discussion

The results of this study indicate that data visualization plays a central role in accelerating the process of data analysis and understanding. Based on interviews with data practitioners, 85% of respondents agreed that good data visualization can significantly reduce the time required to understand patterns and trends in large datasets. This is in line with the theory put forward by Tufte (2001) that simple and effective visualizations can convey information more efficiently than text or numeric tables.

one of the most basic and popular data visualization libraries in Python. It allows the creation of static, animated, and interactive graphs with various customization options. Advanced Customization: Users can modify various elements of the graph, such as labels, titles, scales, legends, colors, and line styles.

Compatibility with Numpy and Pandas: Matplotlib works well with data structures such as Numpy arrays and Pandas DataFrames, which makes it easy to manipulate data before visualizing it.

3D extensions: Matplotlib also supports 3D visualization using the mplot3d module.

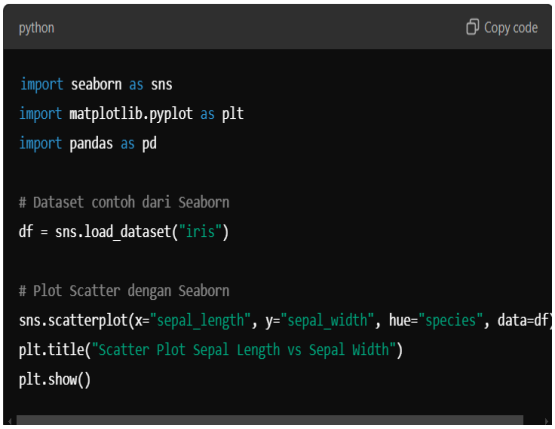


```
import matplotlib.pyplot as plt
import numpy as np

x = np.linspace(0, 10, 100)
y = np.sin(x)

plt.plot(x, y)
plt.title("Grafik Sine Wave")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()
```

Fig 1. Matplotlib



```
python
Copy code
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

# Dataset contoh dari Seaborn
df = sns.load_dataset("iris")

# Plot Scatter dengan Seaborn
sns.scatterplot(x="sepal_length", y="sepal_width", hue="species", data=df)
plt.title("Scatter Plot Sepal Length vs Sepal Width")
plt.show()
```

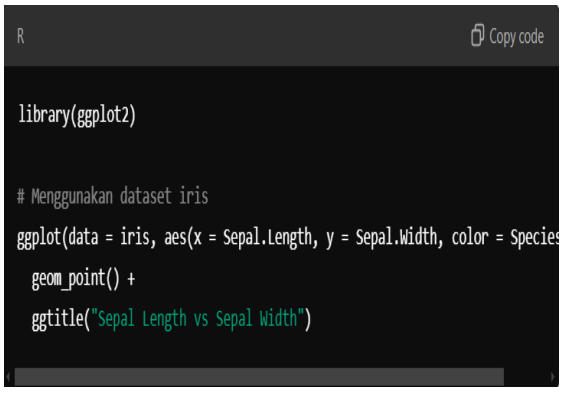
Fig 2. Seaborrn

```
python
Copy code
import plotly.express as px

df = px.data.iris()

fig = px.scatter(df, x="sepal_width", y="sepal_length", color="species"
fig.show()
```

Fig 3. ggplot2



```
R
Copy code
library(ggplot2)

# Menggunakan dataset iris
ggplot(data = iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species
  geom_point() +
  ggtitle("Sepal Length vs Sepal Width")
```

Fig 4. Layered Graphing

Seaborn

- Advanced Statistical Graphs: Seaborn provides tools for creating heatmaps, distribution plots, boxplots, violin plots, pair plots, and more.
- Attractive Color Palette: Seaborn's default color palette is very attractive and easy to use, allowing users to create aesthetically pleasing charts without much customization.
- Integration with Pandas: Seaborn supports Pandas DataFrame, which makes it easier for users to access and visualize data.
- Relationship Visualization: Seaborn supports visualization of relationships between variables such as scatter plots with grouping using color or size**.**

Plotly:

- Interactive Charts: Plotly offers a variety of interactive charts, such as scatter plots, line charts, geographic maps, and heatmaps.
- Web Interface: Plotly output is often displayed in HTML format, allowing easy integration with web apps or interactive dashboards.
- 3D Charts and Maps: Plotly allows the creation of 3D charts and geographic maps, which are useful for spatial data analysis.
- Integration with Dash: Plotly has a framework called Dash, which makes it easy to create data-driven dashboard applications**.**

Layered Graphing: ggplot2 allows for layered graphing using layers, such as data layers, geometric layers (e.g. scatter, line), and scale layers. Customizable Visual Aesthetics: Colors, shapes, and sizes can be freely controlled, allowing for more sophisticated visualizations. Strong Community and Documentation: ggplot2 has a large community support and is well documented.

Conslusion

This study reveals that data visualization plays a key role in facilitating the understanding and analysis of complex data. By transforming raw data into graphical representations, data visualization helps identify hidden patterns, trends, and anomalies, accelerates decision-making, and improves communication between technical and non-technical parties.

The results show that the effectiveness of visualization is greatly influenced by the selection of techniques that are appropriate to the characteristics of the data and the purpose of the analysis. Line charts and heat maps have proven to be very effective in showing trends and relationships between variables, while less appropriate visualizations, such as pie charts in the context of big data, tend to cause misunderstanding or bias.

In addition, visualization tools such as Tableau and Power BI make it easy for users to create interactive visualizations, which can increase user engagement and understanding. However, greater flexibility is offered by D3.js for experienced users who need custom visualizations.

Although data visualization offers many benefits, challenges such as misperceptions and difficulties in handling big data remain. Therefore, greater attention is needed in the selection of scales, colors, and visualization techniques to avoid bias and provide accurate information.

Overall, this study confirms that data visualization is not just a supporting tool in analysis, but a core element that can improve the efficiency and quality of data-based decision-making

processes. The proper use of visualization can provide significant added value to various sectors, especially in the era of ever-growing big data.

Author detail

[1]Universitas Cendekia Mitra Indonesia

Received: 10 Agustus 2024 Accepted: 10 Oktober 2024

Published online: 14 Oktober 2024

References

Bertin, J. (1983). Semiology of Graphics: Symbols, Image, Relations. University of Wisconsin Press.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann.

Cleveland, W. S., & McGill, R. (1985). "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical User Interfaces". Journal of the American Statistical Association, 79(387), 531-554.

Few, S. (2013). Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly Media.

Heer, J., Bostock, M., & Ogievetsky, V. (2010). "A Tour through the Visualization Zoo". Communications of the ACM, 53(6), 59-67.

Keim, D. A., Mansmann, F., & Schneidewind, J. (2013). "Challenges and Opportunities in Visual Analytics". In Visual Analytics: Challenges and Opportunities. Springer.

Kirk, A. (2016). Data Visualisation: A Handbook for Data Driven Design. SAGE Publications.

McCandless, D. (2010). Information is Beautiful. Collins.

Tufte, E. R. (2001). The Visual Display of Quantitative Information. Graphics Press.

Ware, C. (2004). Information Visualization: Perception for Design. Morgan Kaufmann.

Catatan Penerbit

Borneo Novelty Publishing tetap netral sehubungan dengan klaim yurisdiksi dalam peta yang diterbitkan dan afiliasi kelembagaan.